Why we need large trials: The fragility index

PJ Devereaux, MD, PhD
McMaster University, HHS, PHRI
2 hypothetical RCTs

- Evaluating new treatments to prevent MI
- Placebo controlled
- Outcome MI
- Identical methodology
  - concealed, blinded, complete f/u, ITT
    - high quality
<table>
<thead>
<tr>
<th>Trial 1</th>
<th>Tx A (n=100)</th>
<th>Placebo (n=100)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI</td>
<td>1</td>
<td>9</td>
<td>0.02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trial 2</th>
<th>Tx B (n=4000)</th>
<th>Placebo (n=4000)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI</td>
<td>200</td>
<td>250</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Question

• What best characterizes your belief in whether there is a real effect of TxA and TxB?
  • 1. Similar for both trials
  • 2. Substantially more likely for Tx A
  • 3. Modestly more likely for Tx A
  • 4. Substantially more likely for Tx B
  • 5. Modestly more likely for Tx B
<table>
<thead>
<tr>
<th>Trial 1</th>
<th>Tx A (n=100)</th>
<th>Placebo (n=100)</th>
<th>RRR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI</td>
<td>1</td>
<td>9</td>
<td>90% (23-100)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trial 2</th>
<th>Tx B (n=4000)</th>
<th>Placebo (n=4000)</th>
<th>RRR 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI</td>
<td>200</td>
<td>250</td>
<td>20% (5-37)</td>
</tr>
</tbody>
</table>
Question

• What best characterizes your belief in whether there is a real effect of TxA and TxB?
  • 1. Similar for both trials
  • 2. Substantially more likely for Tx A
  • 3. Modestly more likely for Tx A
  • 4. Substantially more likely for Tx B
  • 5. Modestly more likely for Tx B
Small variation in hypothetical trials

• Add 2 events to treatment groups
  – what would happen to p values?
    • 1st trial of 200 patients P value 0.13 (i.e., fragile p value)
    • 2nd trial of 8000 patients P value remains 0.02

• Fragility Index (FI)
  – minimum number of patients required to switch from non-event to event in one group to reverse statistical significance
    • 1st trial FI – 1
    • 2nd trial FI - 9
<table>
<thead>
<tr>
<th>Trial 1</th>
<th>Tx A</th>
<th>Placebo</th>
<th>P value</th>
<th>RRR</th>
<th>FI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n=100)</td>
<td>(n=100)</td>
<td></td>
<td>(95% CI)</td>
<td></td>
</tr>
<tr>
<td>MI</td>
<td>1</td>
<td>9</td>
<td>0.02</td>
<td>90%</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(23-100)</td>
<td></td>
</tr>
<tr>
<td>Trial 1</td>
<td>Tx B</td>
<td>Placebo</td>
<td>P value</td>
<td>RRR</td>
<td>FI</td>
</tr>
<tr>
<td></td>
<td>(n=4000)</td>
<td>(n=4000)</td>
<td></td>
<td>(95% CI)</td>
<td></td>
</tr>
<tr>
<td>MI</td>
<td>200</td>
<td>250</td>
<td>0.02</td>
<td>20%</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(5-37)</td>
<td></td>
</tr>
</tbody>
</table>
What best characterizes your belief in whether there is a real effect of TxA and TxB?

1. Similar for both trials
2. Substantially more likely for TxA
3. Modestly more likely for TxA
4. Substantially more likely for TxB
5. Modestly more likely for TxB
Fragility of trial results

- Highly cited studies in leading journals
  - not uncommonly contradicted (16%)
  - demonstrated to have exaggerated effects (16%)
  - only identified factor explaining these findings
    - outcome in initial trial had a small sample size
What is goal of randomization?

• Achieve balance of prognosis between treatment groups outside of investigational interventions
Balance of prognosis

• Given 9 independent RF associated with MI
  – prevalence of RF varies from 18-65%
  – RF have substantially larger associates with MI
    • (e.g., current smoker OR 2.87) than realistic drug effects
  – not difficult to understand how effect seen in initial 1st trial
    • may be due to imbalance in RFs despite randomization
  – whereas size of 2nd trial minimizes likelihood of meaningful imbalance in RFs that could explain result
Recent experience

- **Beta-blocker trial**
  - randomized 112 patients
  - 11 deaths
    - 2 beta-blocker group, 9 control group
      - $P = 0.02$

- **POISE Trial**
  - randomized 8351 patients
  - 226 deaths
    - 129 beta-blocker group, 97 placebo group
      - $P = 0.03$
Fragility index in high impact RCTs

- JCE 2014 – 399 RCTs published in high impact journals with statistically significant result
- Median sample size n = 682 (range: 15 - 112,604)
- Median of 112 events (range: 8 - 5,142)
- 53% reported a p-value <0.01
- Median FI 8 (range: 0 - 109)
- 25% had FI ≤3
- FI was < number of patients lost to f/u in 53% of trials
Conclusion

• Zeal to turn folks onto RCTs may have resulted in overconfidence in trial results
• Need large sample sizes to achieve balance of prognosis
  – what defines large is focus of ongoing work
• “Positive” trial results frequently hinge on few events
• Fragility Index may improve trial interpretation